Noise Methods for Flavor Singlet Quantities [*]

Walter Wilcox

Physics Department, Baylor University, Waco TX, USA 76798

Abstract. A discussion of methods for reducing the noise variance of flavor singlet quantities ("disconnected diagrams") in lattice QCD is given. After an introduction, the possible advantage of partitioning the Wilson fermion matrix into disjoint spaces is discussed and a numerical comparison of the variance for three possible partitioning schemes is carried out. The measurement efficiency of lattice operators is examined and shown to be strongly influenced by the Dirac and color partitioning choices. Next, the numerical effects of an automated subtraction algorithm on the noise variance of various disconnected loop matrix elements are examined. It is found that there is a dramatic reduction in the variance of the Wilson point-split electromagnetic currents and that this reduction persists at small quark mass.

1 Introduction

1.1 Motivations

The calculation of flavor singlet quantities, also referred to as disconnected diagrams because the fermion lines are disjoint, is one of the greatest technical challenges left in lattice QCD. Disconnected contributions are present in a wide variety of quantities in strong interaction physics including all baryon form factors, axial operators involving quark spin context (Ellis-Jaffe sum rule), hadronic coupling constants and polarizabilities, the p-N sigma term, and propagation functions for various flavor singlet mesons. Such quantities are also present in deep inelastic structure functions measured on the lattice using the operator product expansion, but are not included because of their difficulty and large Monte Carlo error bars. These types of diagrams are difficult to evaluate because exact extractions require many matrix inversions to measure all the background fermion degrees of freedom (including space-time). Disconnected quark contributions are instead isolated stochastically by a process of applying "noises" to the fermion matrix to project out the desired operator contribution. For an overview of selected aspects of flavor singlet calculations in lattice QCD, see Ref.[1].

1.2 Mathematical Background

Noise methods are based upon projection of the signal using random noise vectors as input. That is, given

---

[*] Talk given at the Interdisciplinary Lattice Workshop, Wuppertal, Aug.22-24.

$$Mx = \eta,$$

where M is the quark matrix, x is the solution vector and η is the noise vector, with

$$< \eta_i >= 0, < \eta_i \eta_j >= \delta_{ij},$$

where one is averaging over the noise vectors, any inverse matrix element, M_{ij}^{-1}, can then be obtained from

$$< \eta_j x_i >= \sum_k M_{ik}^{-1} < \eta_j \eta_k >= M_{ij}^{-1}.$$

We shall consider two techniques for reducing the noise variance in lattice QCD simulations: partitioning[2] and subtraction methods[3]. Partitioning the noise appropriately, which means most generally a zeroing out of some pattern of noise vector elements, but which will specifically be implemented here by using separate noise source vectors in Dirac and color spaces, can lead to significant reductions in the variance. Subtraction methods, which involve forming new matrix operators which have a smaller variance but the same expectation value (i.e. are "unbiased"), can also be of great help. The key is using a perturbative expansion of the quark matrix as the subtraction matrices which, however, are not unbiased in general and require a separate calculation, either analytical or numerical, to remove the bias. We will see that the various lattice operators have dramatically different behaviors under identical partitioning or subtraction treatments. Both of these methods, partitioning and subtraction, will be treated numerically but the hope is that the numerical results will eventually be "explained" by some simple rules based on the structure of the Wilson matrix.

2   Noise Theory

2.1   Variance Evaluations

Let us review the basics of matrix inversion using noise theory. The theoretical expressions for the expectation value and variance (V) of matrices with various noises are given in Ref.[4]. One has that

$$X_{mn} \equiv \frac{1}{L} \sum_{l=1}^{L} \eta_{ml} \eta_{nl}^*. \tag{1}$$

$(m, n = 1, \ldots, N; l = 1, \ldots, L.)$ We have

$$X_{mn} = X_{nm}^*, \tag{2}$$

and the expectation value,

$$< X_{mn} >= \delta_{mn}. \tag{3}$$

By definition the variance is given by

$$V[Tr\{QX\}] \equiv < |\sum_{m,n} q_{mn}X_{nm} - Tr\{Q\}|^2 > . \tag{4}$$

The variance may be evaluated as,

$$V[Tr\{QX\}] = \sum_{m \neq n} (< |X_{nm}|^2 > |q_{mn}|^2 \tag{5}$$

$$+ q_{mn}q_{nm}^* < (X_{mn})^2 >) + \sum_{n} < |X_{nn} - 1|^2 > |q_{nn}|^2.$$

2.2   Real Noises

For a general real noise,

$$< |X_{mn}|^2 >= \frac{1}{L}, \tag{6}$$

$$< (X_{mn})^2 >= \frac{1}{L}, \tag{7}$$

for $m \neq n$ so that

$$V[Tr\{QX_{\text{real}}\}] = \frac{1}{L} \sum_{m \neq n} (|q_{mn}|^2 + q_{mn}q_{nm}^*) \tag{8}$$

$$+ \sum_{n} < |X_{nn} - 1|^2 > |q_{nn}|^2.$$

The case of real $Z(2)$ has Eqs.(6) and (7) holding for $m \neq n$, but also

$$< |X_{nn} - 1|^2 >= 0. \tag{9}$$

This shows that

$$V[Tr\{QX_{Z(2)}\}] \leq V[Tr\{QX_{\text{real}}\}]. \tag{10}$$

Thus, $Z(2)$ noise has the lowest variance of any real noise.

2.3   General Z(N) Noise

For general $Z(N)$ $(N \geq 3)$ noise we have a different situation. One has that

$$< |X_{mn}|^2 >= \frac{1}{L}, \tag{11}$$

$$< (X_{mn})^2 >= 0, \tag{12}$$

for $m \neq n$, and again

$$< |X_{nn} - 1|^2 >= 0. \tag{13}$$

Thus

$$V[Tr\{QX_{Z(N)}\}] = \frac{1}{L} \sum_{m \neq n} |q_{mn}|^2, \tag{14}$$

and the variance relationship of $Z(2)$ and $Z(N)$ is not fixed for a general matrix Q. The reason for the difference in Eqs.(7) and (12) is that the square of an equally weighted distribution of $Z(2)$ elements is not itself uniformly distributed (always 1), whereas the square of a uniformly weighted $Z(N)$ distribution for $N \geq 3$ is also uniformly distributed. However, if the phases of q_{mn} and q_{nm}^* are uncorrelated, then $V[Tr\{QX_{Z(2)}\}] \approx V[Tr\{QX_{Z(N)}\}]$, $(N \geq 3)$ which, we will see, is apparently the case for the operators studied here.

3 Partitioning the Problem

3.1 Basic Idea

By "partitioning" I mean replacing the single noise vector problem,

$$Mx = \eta, \tag{15}$$

which yields a complete output column, $\sum_k M_{ik}^{-1} \eta_k$, from a single input noise vector, η, with a problem

$$Mx^p = \eta^p, p = 1, \ldots P, \tag{16}$$

where the η^p have many zeros corresponding to some partitioning scheme. In this latter case it takes P inverses to produce a complete measurement or sampling of a column of M^{-1}.

For the unpartitioned problem (for $Z(N), N \geq 3$, say)

$$V[Tr\{QX\}] = \frac{1}{L} \sum_{m \neq n} |q_{mn}|^2, \tag{17}$$

$$\equiv \frac{1}{L} N(N-1) < |q|^2 >,$$

where I have defined the average absolute squared off diagonal matrix element, $< |q|^2 >$. For the partitioned problem the total variance includes a sum on p,

$$\sum_{p=1}^{P} V[Tr\{QX_p\}] = \frac{1}{L} \sum_{p=1}^{P} \sum_{m_p \neq n_p} |q_{m_p n_p}|^2, \tag{18}$$

$$\equiv \frac{1}{L} N(\frac{N}{P} - 1) < |q_P|^2 > .$$

In order for this method to pay off in terms of computer time, one needs that

$$\sum_{p=1}^{P} V[Tr\{QX_p\}] \leq \frac{1}{P} V[Tr\{QX\}], \tag{19}$$

$$\Rightarrow < |q_P|^2 > \leq \left(\frac{N-1}{N-P}\right) < |q|^2 > . \tag{20}$$

The goal of partitioning is to avoid some of the large off-diagonal matrix elements so that in spite of doing P times as many inverses, a smaller variance is produced for the same amount of computer time. The spaces partitioned can be space-time, color or Dirac or some combination. I have found that partitioning in Dirac and color spaces can strongly affect the results.

3.2   Simulation Description

I consider all local operators, $\bar{\psi}(x)\Gamma\psi(x)$, as well as point-split versions of the vector and axial vector operators. This means 16 local operators and 8 point-split ones, making a total of 24, which are listed below. For each operator there are both real and imaginary parts, but in each case one may show via the quark propagator identity $S = \gamma_5 S^\dagger \gamma_5$, that only the real *or* the imaginary part of each local or nonlocal operator is nonzero on a given configuration for each space-time point. However, this identity is not respected exactly by noise methods, so the cancellations are actually only approximate configuration by configuration. However, the knowledge that one part is purely noise allows one to simply drop that part in the calculations, thus reducing the variance without biasing the answer.

The operators I consider are:

- Scalar: $Re[\bar{\psi}(x)\psi(x)]$
- Local Vector: $Im[\bar{\psi}(x)\gamma_\mu\psi(x)]$
- Point-Split Vector:
    $\kappa Im[\bar{\psi}(x+a_\mu)(1+\gamma_\mu)U_\mu^\dagger(x)\psi(x) - \bar{\psi}(x)(1-\gamma_\mu)U_\mu(x)\psi(x+a_\mu)]$
- Pseudoscalar: $Re[\bar{\psi}(x)\gamma_5\psi(x)]$
- Local Axial: $Re[\bar{\psi}(x)\gamma_5\gamma_\mu\psi(x)]$
- Point-Split Axial:
    $\kappa Re[\bar{\psi}(x+a_\mu)\gamma_5\gamma_\mu U_\mu^\dagger(x)\psi(x) + \bar{\psi}(x)\gamma_5\gamma_\mu U_\mu(x)\psi(x+a_\mu)]$
- Tensor: $Im[\bar{\psi}(x)\sigma_{\mu\nu}\psi(x)]$

I actually consider the zero momentum version of these operators, summed over both space and time.

The sample noise variance in M quantities x_i is given by the standard expression:

$$V_{noise} = \frac{1}{M-1} \sum_{i=1,M} (x_i - \bar{x})^2 \tag{21}$$

What I concentrate on here are the relative variances between the different methods. Since the squared noise error in a single configuration is given by

$$\sigma^2_{noise} = \frac{V_{noise}}{M},\qquad(22)$$

the ratio of variances gives a direct measure of the multiplicative ratio of noises, and thus the computer time, necessary to achieve the same noise error. However, the variance itself does not take into account the extra P inverses done when problem is partitioned. In order to measure the relative efficiency of different partitionings, I form what I call pseudo-efficiencies ratios ("PE"), which are defined by

$$\mathrm{PE}(\frac{\mathrm{method1}}{\mathrm{method2}}) \equiv \frac{P_{method1}(V_{noise})^{method1}}{P_{method1}(V_{noise})^{method2}},\qquad(23)$$

where P_{method} are the number of partitions required by the method. I refer to these ratios as "pseudo" efficiencies since I do a fixed number of iterations for all of the operators I consider. It could very well be that different methods will require significantly different numbers of iterations of conjugate-gradient or minimum residual for the same level of accuracy. One desires to find the lowest PE ratio for a given operator. [1]

I display PE ratio results for Wilson fermions in a $16^3 \times 24$, $\beta = 6.0$ lattice with $\kappa = 0.148$ in Table 1, which follows on the next page. (Part of this Table also appeared in Ref.[2].) I will examine 3 partitionings:

- $Z(2)$ unpartitioned ("$P = 1\ Z(2)$");
- $Z(2)$ Dirac partitioned ("$P = 4\ Z(2)$");
- $Z(2)$ Dirac and color partitioned ("$P = 12\ Z(2)$").

My Dirac gamma matrix representation is:

$$\gamma_i = \begin{pmatrix} 0 & \sigma_i \\ \sigma_i & 0 \end{pmatrix}, \gamma_4 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \gamma_5 = \gamma_1\gamma_2\gamma_3\gamma_4 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}.\qquad(24)$$

In Table 1 I list the relative PEs of the two partitioned methods relative to the unpartitioned case. Referring to the above list of the real or imaginary parts of operators my notation here is as follows: "Scalar" stands for the operator $\bar{\psi}\psi$, "Local Vector 1" for example stands for the operator $\bar{\psi}\gamma_1\psi$, "P-S Vector 1" stands for the 1 component of the point split vector current, "pseudoscalar" stands for $\bar{\psi}\gamma_5\psi$, "Local Axial 1" stands for the operator $\bar{\psi}\gamma_5\gamma_1\psi$, "P-S Axial 1" stands for the point split axial 1 component, and "Tensor 41" stands for example for the operator $\bar{\psi}\sigma_{41}\psi$.

---

[1] Of course an evaluation via N partitioning on an $N \times N$ matrix yields a PE numerator factor of zero relative to other methods since the variance is exactly zero in this case. This is of course prohibitively expensive, but it suggests that more efficient partitionings are possible for large computer budgets. Thanks to M. Peardon for bringing this point out.

There are extremely large variations in the behaviors of the operators listed in Table 1 under identical partitionings. Of the partitionings considered it is most efficient to calculate scalar and vector operators with an unpartitioned simulation. On the other hand, it is far more efficient to calculate the pseudoscalar in a Dirac and color partitioned manner. Notice the entries for the 1,2 components of the axial current (both local and point split) do not behave like the 3,4 components under pure Dirac partitioning, but they do when Dirac and color partitionings are combined. Four of the tensor operators respond best to a pure Dirac partitioning, while the other two prefer a partitioning in Dirac and color spaces combined. The ratio of the largest to the smallest entry in the right hand column is about 800!

As pointed out in Section 2, the variance of $Z(2)$ and $Z(N)$ ($N \geq 3$) noises are in general different. For this reason I also investigated partitioning using $Z(4)$ as well as volume (gauge variant) noises, but there do not seem to be large factors to be gained relative to the $Z(2)$ case. The SESAM collaboration also has seen the efficacy of partitioning (in Dirac space) for axial operators[5].

Table 1. The pseudoefficiency (PE) ratios associated with the methods indicated.

Operator	PE($\frac{\text{P=4 Z(2)}}{\text{P=1 Z(2)}}$)	PE($\frac{\text{P=12 Z(2)}}{\text{P=1 Z(2)}}$)
Scalar	2.83 ± 0.47	10.9 ± 2.5
Local Vector 1	2.38 ± 0.65	8.71 ± 1.8
Local Vector 2	2.50 ± 0.53	12.1 ± 2.9
Local Vector 3	3.60 ± 1.00	11.4 ± 2.4
Local Vector 4	3.41 ± 0.60	16.3 ± 3.4
P-S Vector 1	2.63 ± 0.56	9.94 ± 2.2
P-S Vector 2	2.27 ± 0.44	11.0 ± 2.3
P-S Vector 3	3.52 ± 0.74	11.4 ± 1.5
P-S Vector 4	3.87 ± 0.49	15.5 ± 4.2
Pseudoscalar	0.698 ± 0.15	0.0201 ± 0.0043
Local Axial 1	0.114 ± 0.021	0.144 ± 0.029
Local Axial 2	0.126 ± 0.020	0.146 ± 0.037
Local Axial 3	1.13 ± 0.19	0.162 ± 0.038
Local Axial 4	2.26 ± 0.24	0.187 ± 0.035
P-S Axial 1	0.167 ± 0.040	0.151 ± 0.018
P-S Axial 2	0.110 ± 0.032	0.114 ± 0.028
P-S Axial 3	1.67 ± 0.35	0.186 ± 0.049
P-S Axial 4	1.85 ± 0.21	0.238 ± 0.036
Tensor 41	1.07 ± 0.30	0.295 ± 0.049
Tensor 42	0.345 ± 0.076	0.0889 ± 0.011
Tensor 43	1.32 ± 0.43	0.398 ± 0.13
Tensor 12	1.12 ± 0.25	0.376 ± 0.066
Tensor 13	0.116 ± 0.024	0.363 ± 0.053
Tensor 23	0.0314 ± 0.0058	0.0751 ± 0.016

4   Perturbative Noise Subtraction

4.1   Description of Algorithm

Consider $\tilde{Q}$ such that

$$< Tr\{\tilde{Q}X\} >= 0. \qquad (25)$$

One can then form

$$< Tr\{(Q - \tilde{Q})X\} >=< Tr\{QX\} > . \qquad (26)$$

However,

$$V[Tr\{(Q - \tilde{Q})X\}] \neq V[Tr\{QX\}]. \qquad (27)$$

As we have seen for $Z(N)$ ($N \geq 2$), the variance comes exclusively from off diagonal entries. So, the trick is to try to find matrices $\tilde{Q}$ which are traceless (so they do not affect the expectation value) but which mimic the off-diagonal part of Q as much as possible to reduce the variance.

The natural choice is simply to choose as $\tilde{Q}$ the *perturbative* expansion of the quark matrix. Formally, one has ($I, J = \{x, a, \alpha\}$)

$$(M^{-1})_{IJ} = \frac{1}{\delta_{IJ} - \kappa P_{IJ}}, \qquad (28)$$

where

$$P_{IJ} = \sum_{\mu}[(1 + \gamma_\mu)U_\mu(x)\delta_{x,y-a_\mu} + (1 - \gamma_\mu)U_\mu^\dagger(x - a_\mu)\delta_{x,y+a_\mu}]. \qquad (29)$$

Expanding this in κ gives the perturbative (or hopping parameter) expansion,

$$M_p^{-1} = I + \kappa P + \kappa^2 P^2 + \kappa^3 P^3 + \cdots . \qquad (30)$$

One constructs $< \eta_j(M_p^{-1})_{ik}\eta_k >$ and subtracts it from $< \eta_j M_{ik}^{-1}\eta_k >$, where η is the noise vector. Constructing $< \eta_j(M_p^{-1})_{ik}\eta_k >$ is an iterative process and is easy to code and extend to higher powers on the computer. I will iterate up to 10th order in κ.

One can insert coefficients in front of the various terms and vary them to find the minimum in the variance, but such coefficients are seen to take on their perturbative value, at least for high order expansions[6]. However, see also Ref.[7] where in low orders there is apparently an advantage to this procedure. Interestingly, significant subtraction improvements occur in some operators even in 0th order (point split vectors and two tensor operators.)

For a given operator, $\mathcal{O}$, the matrix $\mathcal{O}M_p^{-1}$ encountered in the context of $< \bar{\psi}\mathcal{O}\psi >= -Tr(\mathcal{O}M^{-1})$ is not traceless. In other words, one must re-add

the perturbative trace, subtracted earlier, to get the full, unbiased answer. How does one calculate the perturbative part? The exact way is of course to explicitly construct all the gauge invariant paths (up to a given κ order) for a given operator. Another approach is to subject the perturbative contribution to a separate Monte Carlo estimation. This is the approach taken here. Local operators require perturbative corrections starting at 4th order (except a trivial correction for $\bar{\psi}\psi$ at zeroth order) and point split ones require corrections starting instead at 3rd order. Because one is removing the bias (perturbative trace) by a statistical method, I refer to this as a "statistically unbiased" method. Some efficiency considerations in carrying out this procedure will be discussed in Section 5. Other versions of subtraction methods in the context of lattice evaluations of disconnected diagrams may be found in Refs.[8] and [9].

4.2   Numerical Results

I am carrying out this numerical investigation in an unpartitioned sense ($P = 1$). The operators which respond best to this partitioning, as discussed previously, are the scalar and local and point-split vector currents, and attention will be limited here to these cases. The effect of combining the partition and subtraction methods has not yet been investigated. I show the ratio of unsubtracted variance divided by subtracted variance, $\mathbf{V}_{\mathrm{unsub}}/\mathbf{V}_{\mathrm{sub}}$ in Figs. 1 and 2. Factors larger than one give the multiplicative gain in computer time one is achieving. The lattices are again Wilson $16^3 \times 24$, $\beta = 6.0$.

Notice the approximate linear rise in the variance ratio as a function of subtraction order for the point-split vector charge density at both $\kappa = 0.148$ and 0.152, Figs. 1 and 2 respectively. Also notice that even at $S = 0$ (subtracting the Kronecker delta) there is a reduction in the variance. The slope of the subtraction graph at $\kappa = 0.148$ is about 3.5; the slope at $\kappa = 0.152$ is reduced to a little under 3.0. Although I do not show the results here, the same linear behavior is evident in $\bar{\psi}\psi$ and the local vector operators although their slopes are considerably smaller.

My final results are summarized in Fig. 3, which gives the reduction in the variance in the scalar and vector operators after a 10th order subtraction has been made at $\kappa = 0.148$. It is not known why the point split vector current responds the best to subtraction. The 10th order point-split vector, local vector, and scalar variance ratios change from ~ 35, ~ 12, and ~ 10 at $\kappa = 0.148$, to $\sim 25 \sim 10$, and ~ 5 at $\kappa = 0.152$, respectively. These are all zero momentum operators. Although the results are not shown, I have found essentially identical results to the above for momentum transformed data, necessary for disconnected form factors. Perturbative subtraction methods will thus be extremely useful in lattice evaluations of nucleon strangeness form factors using the point split (conserved) form of the vector current.

Fig. 1. Effect of the level of perturbative subtraction, up to tenth order in κ, on the ratio of unsubtracted divided by subtracted noise variance for the zero momentum point-split (conserved) charge density operator, J_4, at $\kappa = 0.148$.

Fig. 2. Same as Fig. 1 but for $\kappa = 0.152$.

Fig. 3. Graphical presentation of the unsubtracted variance divided by subtracted variance of 9 different lattice operators after tenth order subtraction at $\kappa = 0.148$ for Wilson fermions. (Abbreviations used for operators: S=scalar; L VEC = local vector; P-S VEC = point split vector).

5   Efficiency Considerations

5.1   Fixed Noise Case

Let me close with some simple observations regarding statistical errors in flavor singlet Monte Carlo simulations. There are two sources of variance in such simulations: gauge configuration and noise. Given N configurations and M noises per configuration, the final error bar on a given operator is given by

$$\sigma = \sqrt{\frac{V_{noise}}{NM} + \frac{V_{gauge}}{N}},\tag{31}$$

where V_{guage} and V_{noise} are the gauge configuration and noise variances. For fixed NM (total number of noises), it is clear that Eq.(31) is minimized by taking $M = 1$. Thus, in this situation it is best to use a single noise per configuration. This simple result is modified by real world considerations of overheads. For example, if one assumes that there is an overhead associated with generating configurations and fixes instead the total amount of computer time,

$$T = NM + G_N N,\tag{32}$$

where G_N is the appropriately scaled configuration generation time overhead, then one finds instead that

$$M = \frac{S_{noise}}{S_{gauge}}\sqrt{G_N},\qquad(33)$$

is the best choice. Note that the ratio S_{noise}/S_{gauge} can have a wide range of values for various operators, and one is no longer guaranteed that $M = 1$ is optimal.

5.2   Fixed Configuration Case

Another common real world situation is where N, the number of configurations, is fixed. In the context of the perturbative subtraction algorithm, one should now maximize the number of effective noises for a given computer budget. The effective number of noises is

$$M_{eff} = M(X + S\Delta s),\qquad(34)$$

where M_{eff} replaces M in Eq.(31). (M retains its meaning as the actual number of gauge field noises.) S is the subtraction order, ranging from 0 to 10 in Figs. 1 and 2, and Δs is the slope. X is the factor one obtains from this method at $S = 0$, without extra overhead. (One sees in Figs. 1 and 2 that this factor is about 2 for the point split charge operator.) I am assuming that the reduction in the variance is approximately linear in subtraction order S. S is imagined to be a continuously variable quantity. The total time per gauge field is

$$T_N = (MT_M + ST_S),\qquad(35)$$

which is kept fixed as M_{eff} is varied. T_M is the noise time overhead and T_S is the subtraction time overhead. The optimum choices for S and M are now

$$S = \frac{T_N}{2T_S} - \frac{X}{2\Delta s},\qquad(36)$$

$$M = \frac{T_N}{2T_M} + \frac{XT_S}{2\Delta sT_M},\qquad(37)$$

resulting in

$$M_{eff} = \frac{\Delta sT_N^2}{4T_ST_M} + \frac{T_NX}{2T_M} - \frac{X^2T_S}{4\Delta sT_M}.\qquad(38)$$

The interesting aspect of this last result is that the effective number of noises, M_{eff}, is now quadratic in T_N. This is a consequence of our observation that the slopes in Figs. 1 and 2 are approximately linear in S. The immediate implication is that for large T_N the noise error bar can be made to vanish

like the inverse of the simulation time rather than as the usual inverse square root, at least in the range of the existing linear behavior.

These equations are also helpful when one has an exact analytical representation of the trace of the perturbative series up to some order, S_{exact}, making T_S zero. Then, by comparing the M_{eff} values in the two cases, one may find the lowest value of S, S_{lowest}, such that $M_{eff}^{lowest} \geq M_{eff}^{exact}$. For example, when $X \approx 0$ and for common values of T_N and T_M in the two simulations, one has

$$S_{lowest} = 2S_{exact}. \qquad (39)$$

That is, for the extra Monte Carlo overhead to pay off, one must attempt to subtract to at least twice as high an order in κ as the exact evaluation. Since subtraction is always exact for nonlocal operators up to second order and for local ones up to third order, making $S_{lowest} \geq 6$ will usually result in a more efficient simulation than the default exact one.

6   Summary

We have seen that significant savings in computer resources may be obtained by partitioning the Wilson matrix appropriately. An efficient partitioning reduces the variance of an operator by leaving out the largest off-diagonal matrix elements of the quark propagator so that in spite of having to do more inversions, a smaller variance is produced for the same amount of computer time. A numerical investigation in Dirac and color spaces revealed efficient partitionings for 24 local and nonlocal operators summarized near the end of Section 3.

We have also seen that large time savings are possible using subtraction methods for selected operators in the context of unpartitioned noise simulations. This method was shown to be effective for the scalar and local vector currents, but most effective for the point-split vector currents. Since perturbative subtraction is based on the hopping parameter expansion of the quark propagator, such methods can become less effective at lower quark masses, although we found the variance reduction was still quite significant for the point-split vector operator at $\kappa = 0.152$. Similar methods can be devised for other operators (axial, pseudoscalar, tensor) by implementing these ideas in the context of Dirac/color partitioned noise methods.

There are still a number of open questions here. The reasons for the strange partitioning patterns found in Section 3 are not known. In addition, the reason why the variance of some operators respond more sensitively to perturbative subtraction than others is obscure. These questions are important because their answers could lead one to better simulation methods. Another question is how far the linear slopes in Figs. 1 and 2 persist at high subtraction orders. Since the perturbative expansion of the Wilson matrix does not converge at small quark mass, the slope of such curves probably

levels off at high enough S. We have seen, however, that before this leveling off occurs the number of effective noises grows quadratically in the simulation time. It was pointed out that this implies that the noise error bar can be made to vanish like the inverse of the simulation time in the range of the existing linear behavior.

7   Acknowledgements

References

1. Güsken, S.: Flavor singlet phenomena in lattice QCD. hep-lat/9906034
2. Wilcox, W. and Lindsay, B.: Disconnected loop noise methods in lattice QCD. Nucl. Phys. **B** (Proc. Suppl.) **63A-C** (1998) 973–975
3. Wilcox, W.: Perturbative subtraction methods. Nucl. Phys. **B** (Proc. Suppl., to appear) hep-lat/9908001
4. Bernardson, S., McCarty, P., and Thron, C.: Monte carlo methods for estimating linear combinations of inverse matrix entries in lattice QCD. Comp. Phys. Comm. **78** (1994) 256–264
5. SESAM Collaboration, J. Viehoff et. al.: Improving stochastic estimator techniques for disconnected diagrams. Nucl. Phys. **B** (Proc. Suppl.) **63** (1998) 269–271
6. Thron, C., Dong, S-J, Liu, K-F, Ying, H-P: Padé-Z2 estimator of determinants. Phys. Rev. **D57** (1998) 1642–1653
7. Mathur, N., Dong, S-J, Liu, K-F and Mukhopadhyay, N.: Proton spin content from lattice QCD. Nucl. Phys. **B** (Proc. Suppl., to appear) hep-lat/9908052
8. Michael C., Foster, M., and McNeile, C.: Flavor singlet pseudoscalar and scalar mesons. Nucl. Phys. **B** (Proc. Suppl., to appear) hep-lat/9909036
9. SESAM collaboration, J. Viehoff et. al.: News on disconnected diagrams. Nucl. Phys. **B** (Proc. Suppl.) **73** (1999) 856–858